Capturing Community Wealth through VTuber Activities

Biligsaikhan Khurtsbayar
2600220460-5

***Abstract*— This study analyzes factors influencing viewer spending on Hololive Japanese female VTuber live streams. Using a dataset of 731 streams from 33 VTubers, we examined stream activity, collaboration, and financial metrics. Analysis shows gaming streams and those during weekday work hours yield higher income, while collaborations negatively impact individual earnings due to audience split. By setting chatter spending more than 2500 yen as wealthy, trained predictive classification model. Due to our small imbalanced dataset, preprocessing included robust scaler and SMOTE. Random Forest model achieved the best performance (F1 score of 0.47, accuracy 0.93). Gaming and collaborator numbers heavily influence the prediction results as key variables. More data is needed for further model refinement and reliable assessment of the research question.**

Keywords: Imbalanced, VTuber, collaboration, gaming, random forest

I. Introduction

As technology advanced, taking its full effect during the COVID-19 pandemic, the nature of entertainment and how we consume it has significantly changed. Shifting away from traditional media, entertainment platforms such as YouTube saw a drastic increase in usage. During this period, YouTube's engagement exploded, increasing by approximately 100 billion minutes and nearly doubling its previous approximation. In this storm of growth, with the combination of their virtual character appeal and entertaining personalities, virtual YouTubers or VTubers managed to take the spotlight and garner the favour of the public [1].

YouTube currently offers three primary content formats: shorts, live streams, and long-form videos (the traditional YouTube format). While shorts and long-form videos generate revenue mainly through ad-based viewership returns, livestreams additionally benefit from YouTube Super Chats monetization. On the other hand, YouTube memberships, not tied to any specific video format, act as a supplementary income stream for creators by giving their subscribers exclusive access to stickers and hidden extra content from their creators.

As the key feature of YouTube live streams, Super Chat enables direct interaction between viewers and creators. Unlike traditional video comment sections, super-chat provides real-time feedback, building stronger connections and growing a more loyal fan base. On top of money, while the viewers incentively remain engaged through reciprocal interactions, such as reading the super-chat, the feature provides creators insight into their personality and activity-based appeal. As seven out of the top ten earners through Super Chats during the pandemic were VTubers, the effectiveness of Vtubers in this content creation approach is apparent [1].

Leading the VTuber market in Japan, operating under Hololive productions, with over 80 million combined subscription count, Cover Corporation manages 86 VTubers. These affiliated VTubers primarily engage in chatting, singing performances, and video games. Additionally, due to the potential influence collaborating has on audience perception, they tend to limit branching out and exclusively operate within their agency, creating a dense collection of audiences [2]. This research focuses on the activities and collaborations of Hololive's Japanese-based female VTubers, analyzing live stream income data to identify behaviours that encourage higher viewer spending or attract wealthier audiences. Based on the collected dataset, the aim is to predict the viewer's wealth with high accuracy.

II. Data Collection

Collected the VTuber livestream earnings data on December 14, 2024, consisting of 27 features and 731 data points.

A. Manually Collected Features

From the official Hololive wiki page, collected the names and debut generation of 33 active female VTubers from Hololive's Japanese main branch. Their YouTube channel IDs were then manually gathered through targeted searches.

1. Channel_name - VTuber name
2. Channel_id - VTuber YouTube channel access id
3. Gen - VTuber's associated Hololive generation (primarily determined by the seniority of their debut)
4. Subscribers - VTuber YouTube channel subscriber count.

B. Tool Retrieved Features

Google YouTube data API retrieves the most recent 30 videos and their associated statistics for each selected Vtuber.

5. Video_name - Extracted video title
6. Video_id - Extracted video access Id
7. Description - Extracted video description
8. Published_at - Published datetime value of the extracted video (For YouTube live streams, the recordings/VOD publish after the live stream ends)
9. Video_start_time - Aired datetime value of the extracted video
10. Video_end_time - End datetime value of the extracted video

11. View_count - Total view count of the extracted video (recording views for live streams)
12. Locale - The speaking language of the VTuber
13. Tags - Associated general topic of the extracted video (such as music, entertainment)

The Chat-downloader library retrieves live stream data for each video.

14. Num_chats - Number of highlighted messages sent to the live stream
15. Num_superchats - The number of paid messages sent to the live stream
16. Val_superchats - The sum of super-chat donations (foreign currency super-chats amounts were converted to yen using the exchange rate from December 14, 2024)
17. Num_memberships - The number of membership subscription messages sent to the live stream

The BeautifulSoup library extracts livestream game name data by parsing the HTML of each video.

18. Game_name - The video game played by the VTuber during the live stream (Hololive VTubers typically play a single video game per stream and start a new stream when switching to a different one)

C. Calculated Features

19. Video_length - The length of the video duration calculated through the difference of video start and end datetime
20. Val_memberships - The total value of membership subscription messages sent to the live stream in yen (membership subscriptions can have varying costs. Due to the inability to access them, set it as the Japanese standard value of 490 yen per subscription)
21. Hashtags - Words following the # symbol in the extracted video description (used by VTubers to set a specified topic for the video, often relating to the activity they are going to participate during it)
22. Ats - Words following the @ symbol in the extracted video description (used by Vtuber to mention the collaborators appearing on the video, often fellow Hololive VTubers)
23. Hashtag_count - Number of # in extracted video description
24. Ats_count - Number of @ in video description
25. Tag_count - Number of the associated general topics of the extracted video
26. Total_val - Total earnings the Vtuber made from the extracted video, excluding video ad revenue
27. Up_time - The total datetime passed from the video publish until the execution of the data collection code

III. Data Cleaning

To start the data preprocessing stage, I checked the dataset for missing or null values. As I collected the data, it was generally clean, with only the game_name feature containing missing values. From the 731 data points, 321 values (43%) were missing in the game_name feature. Further looking into the extracted videos, I determined that these missing values corresponded to non-gaming videos. To resolve this issue, I marked all empty game_name cells as missing and added an additional boolean feature, is_gaming, to indicate whether the video involved gameplay.

I then printed the descriptive statistics of the features and noticed that the minimum values for chats, superchats, and memberships were all zero. Upon investigating, I found that most of these zero values were associated with videos from the VTuber Hoshimachi Suisei. After researching the possible reasons, I discovered that Hoshimachi Suisei is a highly popular VTuber with a successful singing career; her most popular song, "Bibbidiba," was released in early 2024, surpassing 100 million views on YouTube. Following the rapid rise in popularity, during the live stream on May 5, 2024, she announced that she no longer needed superchats and, having achieved financial success, had disabled superchats on her channel. Further, she encouraged her followers to support her by purchasing merchandise or subscribing to her membership instead [3]. Given this context, I considered Hoshimachi Suisei's extracted video data an outlier that could confuse the training process, leading to the removal of all rows associated with her videos from the dataset. Additionally, I excluded all YouTube videos with zero super-chat from the dataset. Since super-chat can be sent only during live streams, videos with zero super-chat are likely regular long-form YouTube videos, resulting in their exclusion, leaving the dataset with 644 data points.

IV. Data Analyis

To analyze the effects of collaboration on activities, an additional feature, is_collaborating, was created, indicated by the ats_count feature having a zero value. Bar graphs were made by grouping the is_gaming and is_collaborating features.

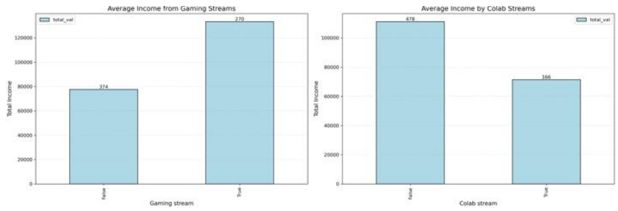

Fig. 1. Average Income per stream grouped by gaming or collaborating

Figure 1 shows that non-collaborative streams generate higher income per video, meaning that collaboration may split the condensed viewer base of the streamer group, resulting in lower overall donations per stream. This observation aligns with previous research findings [3]. As for gaming streams, they appear to generate higher average income than not, requiring further investigation for proper analysis.

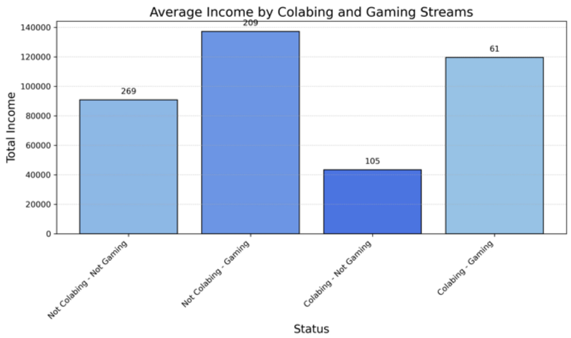

Fig. 2. Average Income per stream grouped by gaming and collaborating

Figure 2 shows that the two variables, is_gaming and is_collaborating, do not cancel each other's insights, meaning our previous observation is now more reliable.

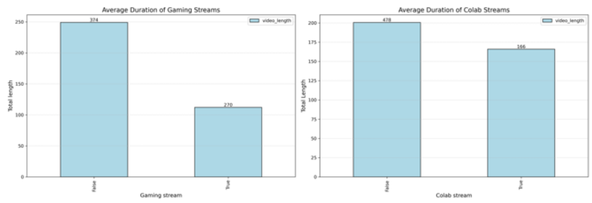

Fig. 3. Average duration of each stream grouped by gaming or collaborating

Finally, Figure 3, when analyzed alongside Figure 1, indicates that stream length is not a significant factor in limiting donation opportunities, as shorter streams do not appear to hinder income generation. So, we can now say conclusively that gaming streams generate more income, whilst collaboration generates less income per stream.

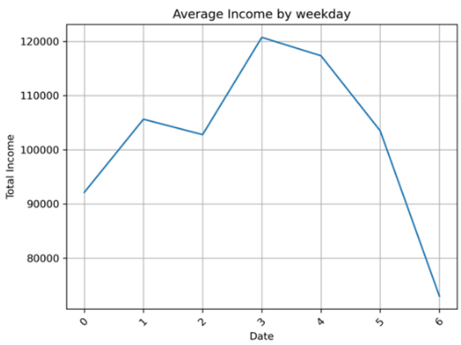

Fig. 4. Average income by day of the week

Using the start time of streams, I created two features: one indicating the weekday the stream occurred and another showing the specific time Vtubers started their streams. From Figure 4, the streams held midweek (Wednesday to Saturday) tend to generate the highest average income.

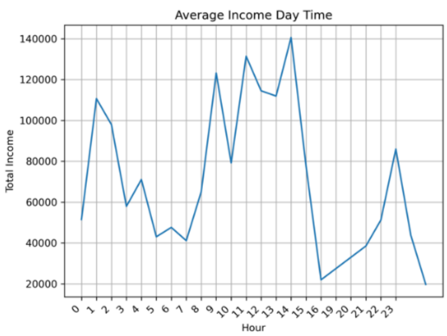

Fig. 5. Average income by time of the week

As for time-based patterns, in Figure 5, there are two noticeable spikes: one during nighttime and another during work hours, indicating that viewers in these periods are more likely to contribute financially.

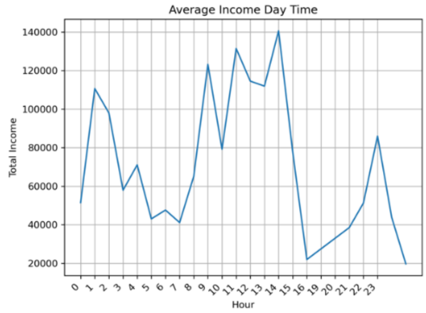

Fig. 6. Average viewer count by day of the week

Figure 6 further supports this observation by showing the average viewer count per hour. Streams during working hours show the least correlation between average income and viewer count, suggesting that individual viewers during this time tend to donate the most. On the other hand, the nighttime income spike is observed due to the sheer number of viewers per stream. To attract high-spending viewers, it seems that Vtubers should focus on streaming during work hours.

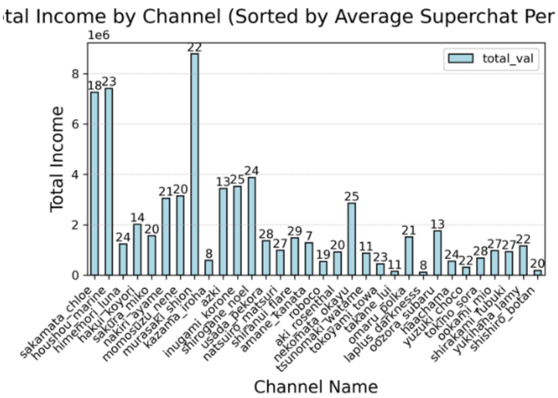

Fig. 7. Average viewer spending per video for each VTuber

To further analyze viewer spending, I created a feature by dividing the total super-chat values by the total number of chats per stream, representing the average spending per chatter for each video. Using this feature, Figures 7 and 8 highlight which VTubers and generations attract the highest average spending viewers.

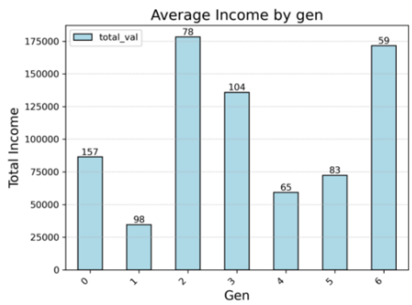

Fig. 8. Average viewer spending per video VTuber in each generation

Initially, I hypothesized that earlier-generation VTubers would have wealthier audiences. However, the data revealed that Generation 6 members (the latest debuts) had the highest average spending viewers, followed by Generation 2. On an individual level, Murasaki Shion and Sakamata Chloe stood out as attracting the most high-spending viewers. This is particularly intriguing, as these two VTubers share a significant overlap in their audience, likely due to their frequent interactions across various social media platforms. In short, the seniority of the Hololive VTubers was not as influential as expected when it comes to viewer wealth.

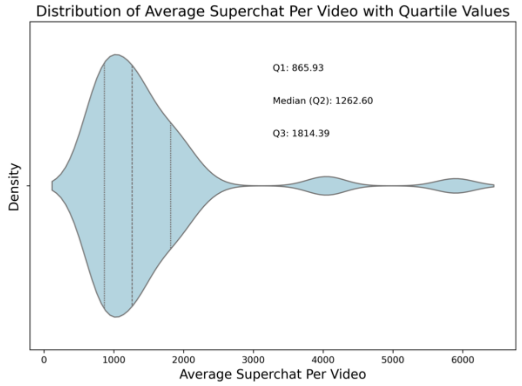

Fig. 9. Average super-chat distribution

In Figure 9, I created a violin plot to visualize the distribution of average spending amounts per chatter for each video. The plot reveals that most chatters spend between 850 and 1250 yen per video, but there are two notable groups of high spenders clustered around 4000 and 6000 yen. These high spenders are the main focus of our prediction model. Based on this analysis, I defined a new feature called chatter_wealth, categorizing a chatter as wealthy if their average spending exceeds 2500 yen, if not being marked as not wealthy. This threshold was chosen due to the clear separation visible in the graph. The chatter_wealth feature will serve as the target variable for training the prediction model.

- Feature engineered features: is_gaming, is_collaborating, hour_of_day, day_of_week_num, average_superchat, chatter_wealth
- Total dataset features: 33 features

V. Training Preprocessing

For my training objective to focus on VTuber activity and collaboration, I first removed irrelevant variables, resulting in a dataset with 16 features. I then categorized these remaining features into textual, categorical, and numerical types.For the textual and categorical features, I applied one-hot encoding to capture the combined effects accurately. For the numerical features, I used RobustScaler to handle potential outliers in the imbalanced dataset. Since RobustScaler uses the interquartile range and median, it preserves the relative influence of outliers, which is key to detecting them [4]. To address the class imbalance seen in Figure 10, I selected SMOTE over SMOTENN, ADASYN, and other advanced variants. Given the relatively clear class separation in this task, I utilized SMOTE as a more computationally efficient approach [5].

Using the preprocessing steps described above, I trained the following models: Random Forest (ensemble learning), Gaussian Naive Bayes (Bayesian), XGBoost (Boosting), and MLP Classifier (FFNN).

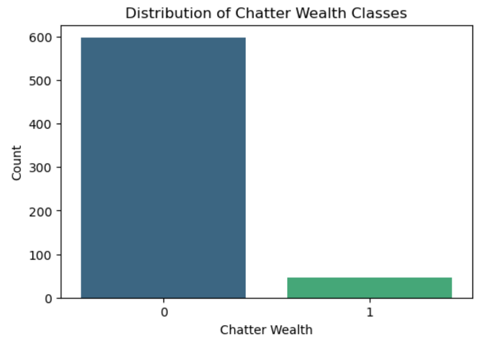

Fig. 10. Chatter wealth class distribution

VI. Training Results and Discussions

Each model mentioned previously was tuned using grid search to find the best parameters for the classification task. With 5-fold cross-validation utilized, the results are also reliable. Due to the heavy class imbalance in the dataset, evaluating the model based on accuracy is less effective. Instead, using the F1 score of the minority class provides a more reliable and insightful measure of performance.

Model name	Minority F1 Score	Accuracy
GaussianNB	0.21	0.58
XGBoost	0.35	0.91
MLP Classifier	0.38	0.92
Random forest	0.47	0.93

a. Classification model results

Fig.11 Prediction model for chatter_wealth result

```
Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.96      0.97      0.96       120
           1       0.50      0.44      0.47         9

    accuracy                           0.93       129
   macro avg       0.73      0.71      0.72       129
weighted avg       0.93      0.93      0.93       129
```

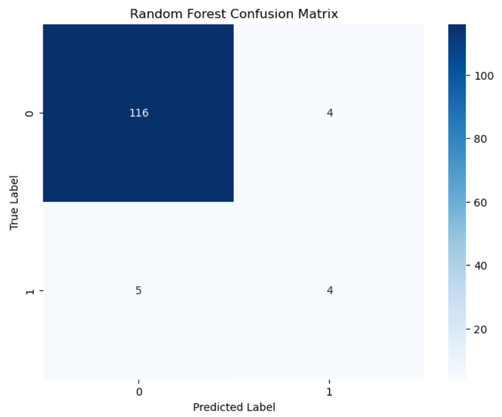

Fig.12 Best mode (Random Forest) classification report

As shown in Figure 11, the Random Forest ensemble learning model achieved an overall accuracy of 0.93 and a minority class F1 score of 0.47, making it the best-performing model in this analysis. From Figure 12, we can see the complete classification report for Random Forest, and it shows the difference between the minority class F1 score of 0.47 and the majority class F1 score of 0.96. The main reason for this drastic change is the heavy imbalance in the dataset. Additionally, due to our small dataset, with only 9 data points in the minority class for evaluation and 36 points used for training (based on a 0.2 train-test split), the model had insufficient data to learn effectively. The most effective way to improve classification performance from this point onwards is to collect more data to better represent the dataset in its entirety.

The MLP classifier performing worse than the Random Forest was an unexpected result. I assumed that an advanced, neural network-based model would outperform other approaches if it could avoid overfitting. However, our specific case was particularly well-suited to ensemble models, which optimize performance by combining multiple models rather than relying on a single strong model. Moreover, the MLP's eagerness to overly focus on the majority class was one of the key reasons behind such underperformance.

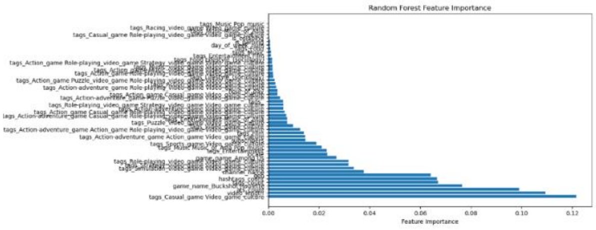

Fig.13 Random Forest feature importance

Additionally, as shown in Figure 13, we can see the top-performing features from the dataset for prediction. The one-hot encoded variables related to gaming categories were the most impactful for the classification task. Additionally, the ats_count or the number of collaborating VTubers was one of the top contributing factors. However, due to the underperformance of the prediction models, we cannot definitively conclude that the behavioural and activity features are the key to attracting a wealthier audience. Still, these features significantly influenced the performance of the best model.

VII. Conclusions

This research attempted to show that specific livestream activities, such as gaming categories and collaborations, can target wealthy audiences and significantly influence the income of Hololive VTubers. Although the results were inconclusive due to the performance of the prediction model, from the data analysis and random forest top feature results, we can tell they were impactful to a certain degree. For future work, to further prove the hypothesis, the dataset needs to be improved in size and minority class. Through the analysis of the audience, we discovered that gen 2 and 6 viewers were comparatively high spending, so the dataset can focus on that region when collecting to resolve the imbalance issue.

References

[1] H. Kim and J. Suh, “Big Data Analysis of ‘VTuber’ Perceptions in South Korea: Insights for the Virtual YouTuber Industry,” Journalism and Media, vol. 5, no. 4, pp. 1723–1748, Nov. 2024, doi: https://doi.org/10.3390/journalmedia5040105.

[2] A. B. Turner, “Streaming as a Virtual Being : The Complex Relationship Between VTubers and Identity,” *DIVA*, 2022. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1676326&dswid=5443 (accessed Jan. 22, 2025).

[3] Yura, “Suisei Turned Superchats Off Because...? (Hoshmachi Suisei / Hololive) [Eng Subs],” YouTube, May 06, 2024. https://www.youtube.com/watch?v=0NmmaFDK1Ko (accessed Jan. 22, 2025).

[4] Y. Singh, “Robust Scaling: Why and How to Use It to Handle Outliers,” *Proclus Academy*, Mar. 22, 2022. https://proclusacademy.com/blog/robust-scaler-outliers/

[5] J. Brownlee, “SMOTE for Imbalanced Classification with Python,” *Machine Learning Mastery*, Jan. 16, 2020. https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/